

La inteligencia artificial fue poco utilizada pero relevante en las revisiones sistemáticas sobre el COVID–19: un estudio metodológico

Juan R Tercero-Hidalgo^{1-3*}, KS Khan¹⁻², A Bueno-Cavanillas¹⁻³, R Fernández-López¹, JF Huete⁴, C Amezcua-Prieto¹⁻³, J Zamora^{2, 5-6}, JM Fernández-Luna⁴

Resumen

Objetivo: Un escenario dinámico como una pandemia requiere la rápida producción de revisiones sistemáticas de calidad, que pueden automatizarse utilizando inteligencia artificial (IA). Se evaluó el uso de herramientas de IA en las revisiones sistemáticas sobre COVID–19.

Diseño del estudio: Tras el registro prospectivo del protocolo del estudio, automatizamos la descarga de todas las revisiones sistemáticas *open-access* sobre COVID–19 en la base de datos *COVID–19 Living Overview of Evidence*, las indexamos en busca de palabras clave relacionadas con la IA y localizamos aquellas que utilizaban herramientas de IA. Comparamos el factor de impacto de sus revistas, las citas por mes recibidas, las cargas de trabajo en *screening*, el tiempo de elaboración (días desde el registro del protocolo hasta el primer *preprint* o envío a una revista) y la evaluación metodológica AMSTAR–2 (máximo, 13 puntos) con un grupo control de revisiones sistemáticas que no usaron IA emparejadas por fecha de publicación.

Resultados: De las 3999 revisiones sobre COVID–19, 28 (0,7%, IC al 95%: 0,47–1,03%) hicieron uso de IA. De media, en comparación con los controles ($n = 64$), las revisiones con IA se publicaron en revistas con mayor factor de impacto (mediana 8,9 vs. 3,5, $P < 0,001$), y examinaron más abstracts por autor (302,2 vs. 140,3, $P = 0,009$) y por estudio incluido (189,0 vs. 365,8, $P < 0,001$), a la vez que inspeccionaron menos *full texts* por autor (5,3 vs. 14,0, $P = 0,005$). No se encontraron diferencias en las citas recibidas (0,5 vs. 0,6, $P = 0,600$), en *full texts* inspeccionados por estudio incluido (3,8 vs. 3,4, $P = 0,481$), en los tiempos de elaboración (74 frente a 123, $P = 0,205$) ni en puntuación AMSTAR–2 (7,5 frente a 6,3, $P = 0,119$).

Conclusión: La IA fue una herramienta infrautilizada en las revisiones sistemáticas sobre COVID–19. Su uso, en comparación con las revisiones sin IA, se asoció con una selección más eficiente de la literatura y un mayor impacto de publicación. Hay cabida para la aplicación de la IA en la automatización de las revisiones sistemáticas.

Keywords

Artificial Intelligence — Systematic Review — COVID–19 — Automation — Research Design — Bibliometrics

¹ Departamento de Medicina Preventiva y Salud Pública, Universidad de Granada

² Instituto Biosanitario de Granada (ibs.GRANADA)

³ CIBER Epidemiología y Salud Pública (CIBERESP)

⁴ Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada

⁵ Unidad de Bioestadística Clínica, Hospital Universitario Ramón y Cajal (IRYCIS)

⁶ Institute for Metabolism and Systems Research, University of Birmingham

* Correspondencia: jrtceroh@gmail.com

Postprint para uso personal,
bajo licencia CC BY-NC-ND 4.0

1. INTRODUCCIÓN

La medicina basada en la evidencia depende de la ágil producción de revisiones sistemáticas para orientar y actualizar la práctica clínica y las políticas sanitarias [1]. Esta es una tarea exigente y costosa, pues requiere que equipos de varios revisores consulten múltiples repositorios y bases de datos, examinen miles de citas y artículos potencialmente relevantes, extraigan los datos pertinentes de los estudios seleccionados y sintetizen sus resultados [2, 3]. En el contexto de la pandemia

de SARS–CoV2/COVID–19, se necesitaban urgentemente métodos para acelerar este laborioso proceso [4, 5].

La elaboración de revisiones sistemáticas requiere seguir procedimientos consistentes y estandarizados para obtener resultados fiables. Sin embargo, la necesidad de acelerar la obtención de resultados durante la pandemia se tradujo en una disminución generalizada de la calidad metodológica de las revisiones [6, 7] y la popularización de las “revisiones rápidas” [8, 9] (que acortan los plazos habituales de producción sacrificando en rigor de la búsqueda, precisión del *screening* o

la extracción de datos y a costa de un mayor riesgo de sesgos). ¿Son estas omisiones inevitables para obtener resultados más rápidos?

En contraste, las soluciones basadas en inteligencia artificial (IA) automatizan partes del flujo de trabajo imitando la resolución humana de problemas (incluyendo el uso de *machine-learning*, procesamiento del lenguaje natural, minería de datos y otros subcampos) [10] para complementar o sustituir los esfuerzos humanos con un riesgo limitado de sesgos [11–13], y se han empleado previamente pero de forma escasa [14] en revisiones sistemáticas con el fin de mejorar el *screening* [15] y la extracción de datos [16, 17]. Su objetivo es acortar los tiempos de producción, permitir un cribado más amplio de la literatura y reducir la carga de trabajo de los revisores sin comprometer la calidad metodológica.

En este trabajo, evaluamos el uso de las herramientas basadas en IA en las revisiones sistemáticas sobre el COVID-19 para determinar empíricamente si, en comparación con las revisiones sobre el COVID-19 sin IA, tuvieron un impacto en la producción, la calidad y la publicación de las revisiones sistemáticas.

2. MATERIALES Y MÉTODOS

Este estudio metodológico [18] se ha elaborado de acuerdo con las directrices PRISMA 2020 [19] (checklist proporcionada como [Material Suplementario 1A](#)), y su protocolo se registró en *Open Science Forum Registries* (DOI [10.17605/OSF.IO/H5DAW](#)) [20] de forma prospectiva.

2.1 Búsqueda y selección de revisiones

Se consideraron para su inclusión todas las revisiones sistemáticas relacionadas con el COVID-19 que pudieran haber hecho uso de cualquier herramienta de IA (*machine-learning*, *deep-learning* o procesamiento del lenguaje natural) para acelerar, mejorar o complementar cualquier aspecto de la realización de la revisión (búsqueda, *screening*, extracción de datos y síntesis). Implementamos un *script* (disponible en el DOI [10.5061/dryad.9kd51c5j6](#)) [21] para procesar todas las referencias bibliográficas registradas en la base de datos *COVID-19 Living Overview of Evidence (L-OVE)* [22] filtrando aquellas clasificadas como “revisión sistemática” entre el 1 de diciembre de 2019 y el 15 de agosto de 2021, y, a continuación, consultar la base de datos *Unpaywall* [23] con cada DOI extraído para obtener un registro JSON con enlaces de descarga. El proceso se repitió 3 veces desde la publicación de nuestro protocolo con el objetivo de reducir la potencial pérdida puntual de revisiones por errores de conexión con dichos servidores (la última búsqueda fue el 17 de agosto de 2021).

Para seleccionar las revisiones que utilizaron IA, elaboramos una lista de palabras clave con alta probabilidad de aparecer en artículos con herramientas de IA ([Material Suplementario 1B](#)). Indexamos cada archivo descargado con el motor de búsqueda *OpenSemanticSearch*, que ejecutamos en una máquina virtual local. Cada archivo que contenía alguna

de nuestras palabras clave fue inspeccionado manualmente de forma independiente por dos autores (JRTH y RFL). Se incluyeron *preprints* y artículos en otros idiomas además del inglés. El único criterio de exclusión aplicado fue la no disponibilidad de los archivos en modalidad *open-access*, debido a la necesidad de evaluar la sección de métodos de cada revisión incluida. Para crear un grupo de comparación con suficiente poder estadístico de revisiones sin IA, por cada revisión incluida se utilizaron los registros obtenidos para seleccionar aleatoriamente 3 controles con la misma fecha de publicación (con un margen de un día si no había suficientes revisiones disponibles en una fecha determinada). Además, localizamos e incluimos en el análisis todas las versiones anteriores de las revisiones categorizadas como *living* o como actualizaciones de otra publicada previamente (*updated*).

2.2 Extracción de datos

Dos autores (JRTH y RFL) extrajeron manualmente los siguientes datos de cada revisión: tipo de revisión (según la descripción de sus autores: estándar, *rapid* o *scoping*, *living* o *updated*); información sobre la financiación recibida y los conflictos de intereses declarados; estado de publicación, factor de impacto en el *Journal Citation Reports (JCR)* de 2020 de la revista que la publica y número de citas recibidas (hasta el 17 de agosto de 2021); número de abstracts inspeccionados, *full texts* revisados y estudios primarios incluidos; número de autores y de revisores que participaron en el *screening*; y fechas de registro del protocolo (si estaba disponible) y de la versión más antigua de la revisión. Para las revisiones *living* y *updated*, se calculó la diferencia de abstracts inspeccionados y estudios incluidos entre cada una de sus versiones y se atribuyó su recuento de citas a la más reciente (para evitar contabilizarlas doblemente). Se utilizó Excel para registrar todas las variables.

Tres autores (JRTH y RFL, asistidos por CAP) evaluaron todas las revisiones con la herramienta AMSTAR-2 de calidad metodológica y riesgo de sesgos [24]. Se excluyeron sus ítems 11-12 y 15, aplicables únicamente a meta-análisis (según lo preestablecido en nuestro protocolo) y se otorgaron 0,5 puntos a las respuestas “*parcialmente sí*” cuando correspondiera, permitiendo así una puntuación máxima de 13 puntos. Para las revisiones *living* y *updated*, sólo se evaluó su versión más reciente (para evitar contabilizarlas doblemente). En las revisiones que incluían tanto ensayos controlados aleatorizados como estudios observacionales, la pregunta 9 (evaluación del riesgo de sesgo de los estudios individuales) se evaluó por separado para cada tipo de estudio. La lista de evaluación utilizada se proporciona como [Material Suplementario 1C](#).

2.3 Síntesis de datos

Se calcularon los ratios de abstracts y *full texts* inspeccionados por autor (como medida de la carga de trabajo) y por estudio incluido (como precisión del *screening*). El número de revisores que participaron en el *screening* se reportó de forma inconsistente entre los estudios y, por tanto, no se utilizó en los cálculos. Se calculó el tiempo de elaboración de

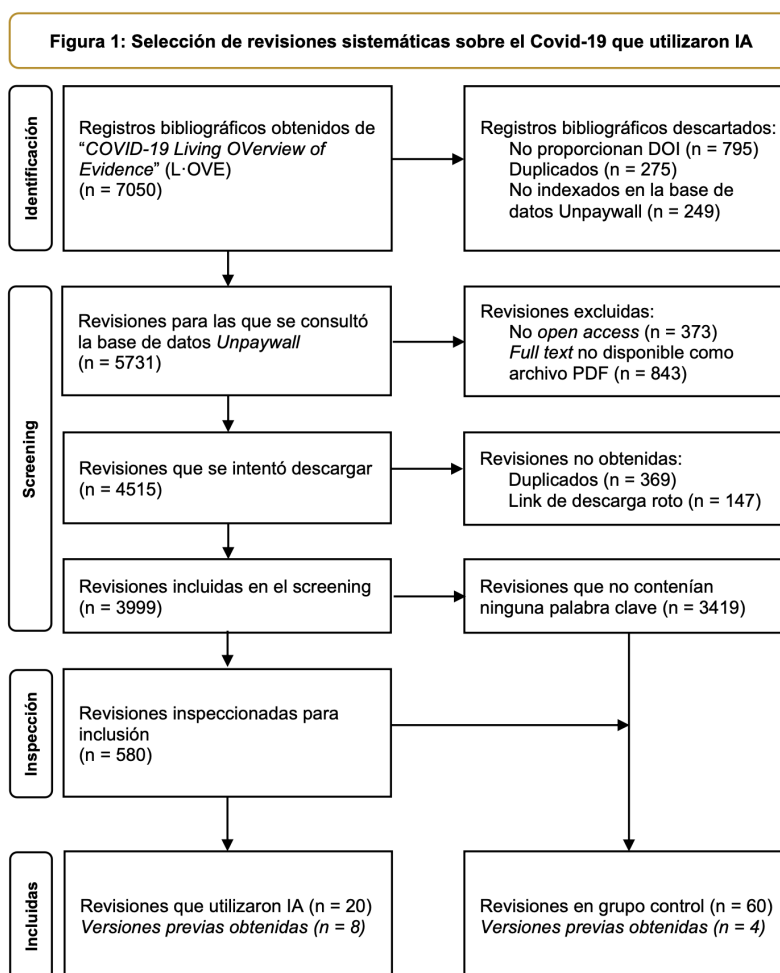


Figura 1 Diagrama de flujo de las revisiones sistemáticas identificadas, cribadas, inspeccionadas para elegibilidad e incluidas en nuestro estudio.

las revisiones con protocolo pre-registrado como la diferencia entre la fecha de su protocolo y la fecha de publicación del primer *preprint* (o la fecha de recepción en la revista, en el caso de los artículos publicados sin *preprint* disponible). Los tiempos de elaboración de las revisiones *living* y *updated* se calcularon como la diferencia entre las fechas de publicación de cada una de sus versiones. Se excluyeron de esta variable las revisiones no pre-registradas debido a la heterogeneidad con que se informaron sus fechas de inicio. Se utilizó la prueba de chi-cuadrado de Pearson para comparar el porcentaje de revisiones rápidas, *living*, que recibieron financiación y que fueron publicadas entre los grupos. El Factor de Impacto JCR de las revistas publicadas, los recuentos de citas recibidas, las cargas de trabajo de la revisión, los tiempos de elaboración y las calificaciones AMSTAR–2 se reportaron como medianas con rangos intercuartílicos (IQR), se representaron mediante diagramas *box-and-whisker*, y se compararon mediante el test de Wilcoxon-Mann-Whitney. Se utilizó el software R (versión 4.0.5) para el cálculo estadístico y GraphPad Prism 9.2.0 para los gráficos. También se proporcionó una descripción narrativa de las revisiones que utilizaron inteligencia artificial, detallando qué partes del proceso de la revisión se automatizaron y

qué software fue utilizado, cómo difirieron las calificaciones de AMSTAR–2 entre ellas, y cómo los autores justificaron o qué impacto atribuyeron al uso de las herramientas basadas en IA.

3. RESULTADOS

3.1 Búsqueda y selección de revisiones

Como se muestra en la *Figura 1*, identificamos 7050 registros bibliográficos de revisiones sistemáticas sobre el COVID–19, descargamos con éxito 3999 e inspeccionamos manualmente 580 que contenían algunas de nuestras palabras clave. Seleccionamos 20 revisiones, de las cuales localizamos 8 versiones anteriores, sumando así 28 revisiones (0,7% del total, intervalo de confianza al 95%: 0,47–1,03%) con uso de IA. De las 60 revisiones seleccionadas como controles según su fecha de publicación, localizamos otras 4 versiones anteriores, lo que hace un total de 64 revisiones sin uso de IA. La lista completa de revisiones sistemáticas seleccionadas se proporciona en formato Excel (*Material Suplementario 2*, sección “*Revisiones incluidas*”) con todas las variables extraídas y el desglose de la herramienta AMSTAR–2 con cada ítem

Tabla 1 Variables extraídas para las revisiones que usaron inteligencia artificial (IA) y los controles

Características	Grupo IA (n = 20)		Controles (n = 60)		Δ	χ^2	Valor P
	n	(%)	n	(%)			
Revisiones tipo <i>rapid</i>	5	(25%)	6	(10%)	15%	2.846	0.092
Revisiones tipo <i>living</i>	5	(25%)	3	(5%)	20%	6.667	0.010
Recibieron financiación	12	(60%)	27	(45%)	15%	1.351	0.245
Publicadas	12	(60%)	48	(80%)	-20%	3.2	0.074
	Mediana	IQR	Mediana	IQR		Wilcoxon W	Valor P
JCR Impact Factor de la revista	9	(4-40)	3	(3-6)		409	<0.001
Citas recibidas por mes	1	(0-13)	1	(0-3)		647	0.600
Abstracts inspeccionados							
por autor	302	(127-804)	140	(44-378)		1126	0.009
por estudio incluido	189	(94-366)	27	(14-64)		1443	<0.001
Full texts inspeccionados							
por autor	5	(4-16)	14	(7-37)		504.5	0.005
por estudio incluido	4	(2-5)	3	(2-6)		883.5	0.481
Tiempo de elaboración (días)	74	(48-118)	123	(53-221)		183.5	0.205
Puntuación AMSTAR-2 (sobre 13)	8	(5-9)	6	(4-8)		740.5	0.119

Se utilizó la prueba de chi-cuadrado de Pearson para comparar las proporciones de revisiones *rapid*, *living*, que recibieron financiación y que fueron publicadas, y el test de Wilcoxon-Mann-Whitney para el resto de comparaciones. Las medianas y rangos intercuartílicos (IQR, expresados como *cuartil 1* – *cuartil 3*) se muestran redondeados al entero más cercano.

Δ : Diferencias absolutas en puntos porcentuales entre las revisiones que usaron IA y las de control. χ^2 : estadístico de la prueba de chi-cuadrado de Pearson. Wilcoxon W: estadístico del test de Wilcoxon-Mann-Whitney (rank-sum test).

evaluado. También se proporciona la lista completa de revisiones inspeccionadas manualmente y finalmente no incluidas en el trabajo (sección “Revisiones excluidas”).

3.2 Descripción de las revisiones incluidas

Las variables extraídas se resumen en la [Tabla 1](#) y pueden visualizarse en la [Figura 2](#). De las 20 revisiones seleccionadas por utilizar IA, hubo 5 *rapid reviews* (25%, con 1 *scoping review* y 1 *rapid evidence map*) y 5 revisiones *living* (25%). Quince revisiones proporcionaron una declaración de conflictos de intereses, de las cuales 12 (60%) declararon haber recibido financiación externa; 12 (60%) habían sido publicadas. De las 60 revisiones del grupo control, hubo 6 *rapid reviews* (10%, con 1 *scoping review*) y 3 revisiones *living* (5%). Cincuenta y siete revisiones proporcionaron una declaración de conflicto de intereses, de las cuales 27 (45%) declararon haber recibido financiación externa; 48 (80%) habían sido publicadas. Los Factores de Impacto JCR y los recuentos de citas recibidas mostraron una alta variabilidad en el grupo IA, principalmente debido a la inclusión de 3 revisiones publicadas en BMJ [25–27], 2 revisiones Cochrane [28, 29] y 1 revisión de la revista Lancet [30]. Además, sólo 10 revisiones en el grupo IA (50%) y 22 en los controles (36%) registraron previamente un protocolo, lo que en conjunto supuso únicamente 44 registros para el cálculo de los tiempos de elaboración.

3.3 Comparación de las revisiones con IA con los controles

El grupo IA incluyó más revisiones *living* que los controles (5/20 vs. 3/60, IC del 95% diferencia absoluta 0,2 a 39,8%, $P = 0,010$), pero no mostró diferencias en revisiones *rapid* (5/20 vs. 6/60, IC del 95% -5,4 a 35,4%, $P = 0,092$), financiación (12/20 vs. 27/60, IC del 95% -9,9 a 39,9%, $P = 0,245$) o estado de publicación (12/20 vs. 48/60, IC del 95% -43,7 a

3,7%, $P = 0,074$). Los factores de impacto JCR en las revisiones publicadas en el grupo IA fueron significativamente mayores que los controles (mediana [IQR]: 8,9 [3,9-39,9] vs. 3,5 [2,6-5,5], $P < 0,001$); en número de citas recibidas no mostraron diferencias (0,5 [0,0-13,5] vs. 0,6 [0,0-2,8], $P = 0,600$).

En cuanto a las mediciones de la carga de trabajo, el grupo IA revisó más abstracts por autor (302,2 [126,7-804,3] frente a 140,3 [43,8-378,2], $P = 0,009$) y por estudio incluido (189,0 [94,1-365,8] frente a 26,9 [13,7-64,1], $P < 0,001$), mientras que inspeccionaron menos *full texts* por autor (5,3 [3,7-16,1] vs. 14,0 [6,5-37,2], $P = 0,005$) y por estudio incluido (3,8 [2,4-5,3] vs. 3,4 [2,0-6,2], $P = 0,481$).

No se observaron diferencias en los tiempos de elaboración de las revisiones preregistradas (74,0 [47,5-117,5] frente a 123,0 [53,0-221,0], $P = 0,205$). Las puntuaciones medias obtenidas en la evaluación metodológica AMSTAR-2 ([Material Suplementario 1C](#)) no fueron significativamente mayores en el grupo IA (7,5 [5,3-9,1] vs. 6,3 [3,9-8,0] puntos sobre 13, $P = 0,119$), mostrando ambos grupos una alta heterogeneidad de resultados, como se puede observar en la [Figura 3](#). En comparación con los controles, las revisiones con IA obtuvieron peores resultados en la pregunta 4 (estrategia de búsqueda de literatura, -12%) y mejores en la pregunta 6 (extracción de datos por duplicado, 35%), mientras que mostraron diferencias mínimas en la pregunta 5 (*screening* por duplicado, 7%). Ambos grupos obtuvieron las puntuaciones más bajas en las preguntas 7 (proporcionar una lista de estudios excluidos) y 10 (informar sobre las fuentes de financiación de los estudios incluidos).

3.4 Descripción narrativa de los usos de la IA

Según la etapa del proceso de revisión en la que se utilizó IA, podemos clasificar las 20 revisiones del grupo IA en tres categorías, como se muestra en la [Tabla 2](#).

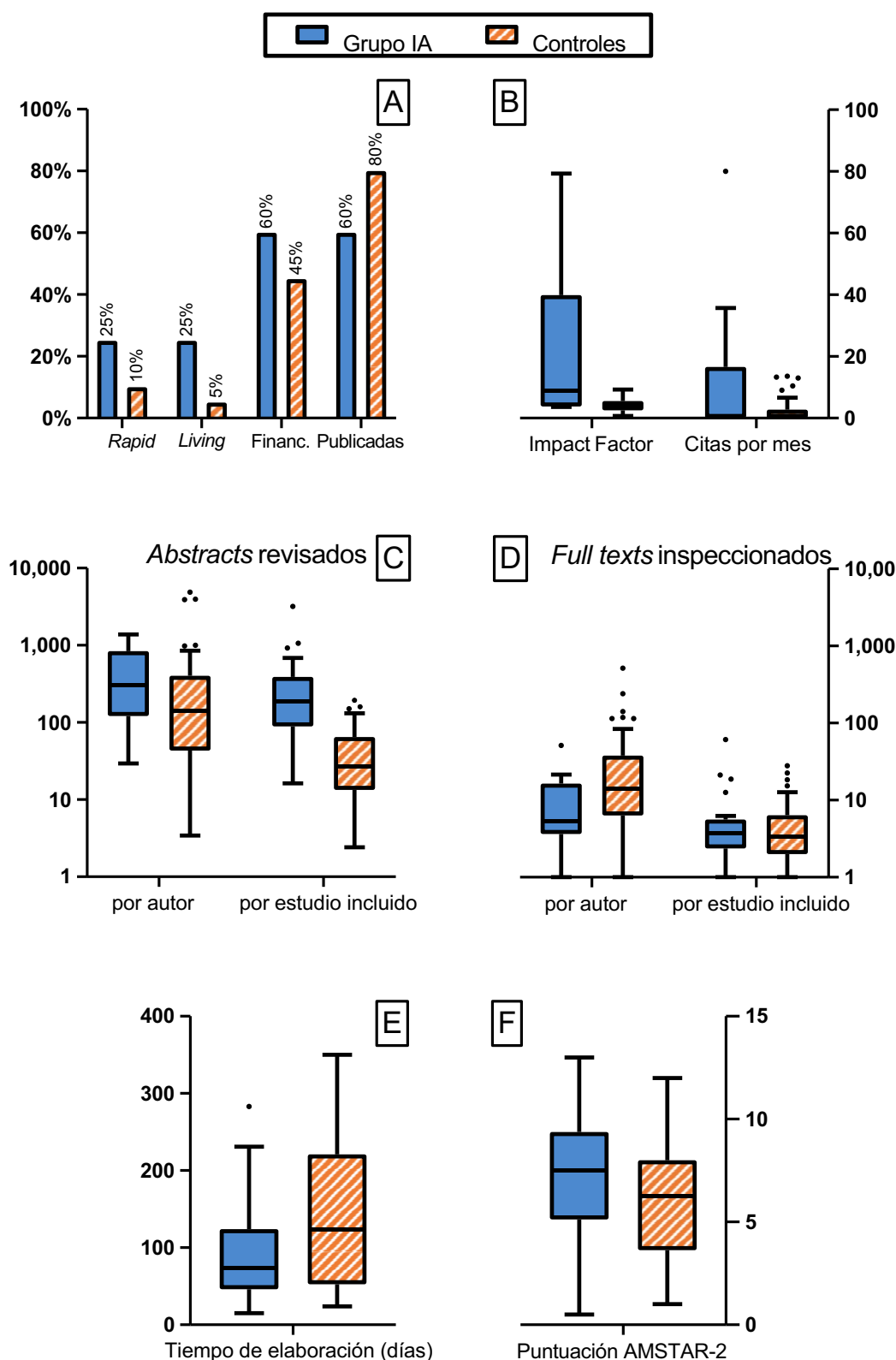


Figura 2 Características de las revisiones incluidas. Diagrama *box-and-whisker* (de “cajas y bigotes”: las cajas encierran los cuartiles Q1–Q3, sus líneas centrales representan la mediana, y los bigotes se extienden hasta los puntos de datos más lejanos dentro del intervalo de 1,5 IQR). El panel A compara la proporción de revisiones *rapid*, *living*, financiadas y publicadas entre los grupos; el panel B presenta los Factores de Impacto JCR 2020 de las revistas y los recuentos de citas por mes de cada grupo; los paneles C y D muestran las mediciones de la carga de trabajo de los autores: abstracts revisados y *full texts* inspeccionados, por autor y por estudio incluido; el panel E muestra los tiempos medios de elaboración (en días) de las revisiones de cada grupo; y el panel F representa sus evaluaciones metodológicas AMSTAR-2.

Tabla 2a Uso de IA: Asistencia en la búsqueda bibliográfica

Ref.	Título	Autor	Revista	Software utilizado	¿Código abierto?
[31]	Prevalence of Gastrointestinal Symptoms and Fecal Viral Shedding in Patients with Coronavirus Disease 2019	Parasa <i>et al.</i>	JAMA Network Open	CORD-19	Parcial
[32]	The influence of comorbidity on the severity of COVID-19 disease: systematic review and analysis	Zaki <i>et al.</i>	preprint	CORD-19 + Okapi BM25	Sí
[33]	The Estimations of the COVID-19 Incubation Period: A Scoping Reviews of the Literature	Zaki <i>et al.</i>	Journal of Infection and Public Health	CORD-19 + BioBERT	Sí
[34]	Ocular toxicity and Hydroxychloroquine: A Rapid Meta-Analysis	Michelson <i>et al.</i>	preprint	GenesisAI	No
[35]	A Systematic Review of the Incubation Period of SARS-CoV-2: The Effects of Age, Biological Sex, and Location on Incubation Period	Daley <i>et al.</i>	preprint	No reportado	No

Tabla 2b Uso de IA: Filtrado de ensayos controlados aleatorizados

Ref.	Título	Autor	Revista	Software utilizado	¿Código abierto?
[36]	Impact of remdesivir on 28 day mortality in hospitalized patients with COVID-19: February 2021 Meta-analysis	Robinson <i>et al.</i>	preprint	RobotSearch	Sí
[37]	Impact of systemic corticosteroids on hospitalized patients with COVID-19: January 2021 Meta-analysis of randomized controlled trials	Robinson <i>et al.</i>	preprint	RobotSearch	Sí
[25]	Prophylaxis against COVID-19: living systematic review and network meta-analysis	Bartoszek <i>et al.</i>	BMJ	RobotSearch	Sí
[26]	Drug treatments for COVID-19: living systematic review and network meta-analysis	Siemieniuk <i>et al.</i>	BMJ	RobotSearch	Sí
[38]	Adverse effects of remdesivir, hydroxychloroquine, and lopinavir/ritonavir when used for COVID-19: systematic review and meta-analysis of randomized trials	Izcovich <i>et al.</i>	preprint	RobotSearch	Sí
[39]	Tocilizumab and sarilumab alone or in combination with corticosteroids for COVID-19: A systematic review and network meta-analysis	Zeraatkar <i>et al.</i>	preprint	RobotSearch	Sí
[40]	Clinical trials in COVID-19 management & prevention: A meta-epidemiological study examining methodological quality	Honarmand <i>et al.</i>	Journal of Clinical Epidemiology	RobotSearch	Sí

Tabla 2c Uso de IA: Automatización del screening

Ref.	Título	Autor	Revista	Software utilizado	¿Código abierto?
[41]	Impacts of school closures on physical and mental health of children and young people: a systematic review	Viner <i>et al.</i>	preprint	EPPI Reviewer	No
[27]	Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal	Wynants <i>et al.</i>	BMJ	EPPI Reviewer	No
[28]	Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection (Review)	Dinnes <i>et al.</i>	Cochrane Database of Systematic Reviews	EPPI Reviewer	No
[29]	Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19	Struyf <i>et al.</i>	Cochrane Database of Systematic Reviews	EPPI Reviewer	No
[42]	Are medical procedures that induce coughing or involve respiratory suctioning associated with increased generation of aerosols and risk of SARS-CoV-2 infection? A rapid systematic review	Wilson <i>et al.</i>	Journal of Hospital Infection	EPPI Reviewer	No
[43]	Risk and Protective Factors in the COVID-19 Pandemic: A Rapid Evidence Map	Elmore <i>et al.</i>	Frontiers in Public Health	SWIFT-Active Screener	No
[44]	Tocilizumab and Systemic Corticosteroids in the Management of COVID-19 Patients: A Systematic Review and Meta-Analysis	Alkofide <i>et al.</i>	International Journal of Infectious Diseases	Abstrackr	Sí
[30]	Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis	Chu <i>et al.</i>	The Lancet	Evidence Prime	No

Tabla 2 Tabla que muestra las diferentes herramientas de inteligencia artificial (IA) que se utilizaron en la elaboración de revisiones sistemáticas sobre COVID-19 (con enlaces a aquellas de código abierto), según su ámbito de aplicación: asistencia en la búsqueda bibliográfica, filtrado de ensayos controlados aleatorizados (ECA) y automatización del screening.

Asistencia en la búsqueda bibliográfica

Tres revisiones [31–33] complementaron sus procedimientos de búsqueda con consultas de tipo “pregunta abierta” en CORD-19 [45], una base de datos sobre el COVID-19 estructurada para facilitar el uso de sistemas de minería de textos y *deep-learning*: Zaki *et al.* [32] utilizaron un repositorio de GitHub basado en el algoritmo de búsqueda *Okapi BM25*; Zaki *et al.* [33] emplearon *BioBERT*, un sistema revisado por pares [46] y de código abierto preentrenado para el análisis de literatura biomédica; y Parasa *et al.* [31] no proporcionaron detalles sobre el motor de búsqueda empleado. Además, Michelson *et al.* [34] utilizaron software privado de la empresa *GenesisAI* para producir un “*rapid meta-analysis*” como prueba de concepto de su producto. Daley *et al.* [35] no proporcionaron ninguna información sobre el software empleado. Sólo 2 revisiones en este subgrupo se encontraban publicadas, y ninguna pre-registró un protocolo. La puntuación media en AMSTAR-2 fue de 3,7/13.

Filtrado de ensayos controlados aleatorizados

Siete artículos [25, 26, 36–40] emplearon *RobotSearch*, un software revisado por pares [47] y de código abierto para identificar, de entre las referencias bibliográficas proporcionadas por el usuario, aquellas correspondientes a ensayos controlados aleatorizados (ECA). Se basa en una red neuronal entrenada con datos de las revisiones de Cochrane y destaca por su facilidad de uso (no requiere instalación) y flexibilidad (ya que permite diferentes niveles de sensibilidad, incluido uno destinado específicamente a revisiones sistemáticas, así como la integración con otros *scripts*).

En nuestra muestra, *RobotSearch* se utilizó particularmente con frecuencia en revisiones *living* o parcialmente automatizadas. Dos de las revisiones que utilizaron *RobotSearch* fueron la de Bartoszko *et al.* [25] un meta-análisis en red sobre la profilaxis del COVID-19, y Siemieniuk *et al.* [26], un *living meta-analysis* de ensayos aleatorizados para guiar las recomendaciones de la Organización Mundial de la Salud (OMS) sobre la terapéutica del COVID-19, de los cuales Izcovich *et al.* [38] y Zeraatkar *et al.* [39] son subestudios separados. Ambos forman parte del proyecto “*BMJ Rapid Recommendations*” y publican sus conclusiones y análisis preliminares en un sitio web en constante actualización. La puntuación media en AMSTAR-2 fue de 7,5/13.

Automatización del *screening*

Encontramos ocho artículos [27–30, 41–44] que hacían uso de procedimientos de *screening* potenciados por IA. Cinco de ellos [27–29, 41, 42] utilizaron *EPPI Reviewer*, una plataforma web (de acceso tipo *shareware*) para asistir en la elaboración de todo tipo de revisiones sistemáticas. Ofrece una gran variedad de funciones, desde la gestión bibliográfica hasta facilitar el trabajo colaborativo, así como la capacidad de sugerir otros estudios relevantes, la agrupación automática de artículos y la minería de textos. En particular, las revisiones incluidas utilizaron su módulo “*SGDClassifier*” para priorizar el *screening* de artículos con más probabilidad de ser incluidos

de acorde con las decisiones previas del usuario. Como resultado, tanto Wynants *et al.* [27] como dos revisiones Cochrane [28, 29] citan una reducción del 80 % en la carga de *screening* gracias a esta herramienta.

Otras dos revisiones utilizaron técnicas de automatización del *screening* englobadas en plataformas de asistencia a la elaboración de revisiones sistemáticas: *SWIFT-Active Screener* [48] en Elmore *et al.* [43], configurado para asegurar un recall mínimo (porcentaje de captura de estudios relevantes) como criterio de detención del *screening*; y *Evidence Prime* de Chu *et al.* [30] para verificar y complementar el *screening* manual. Por último, Alkofide *et al.* [44] utilizaron *Abstrackr*, el único software de código abierto en esta categoría, que utiliza el feedback de los artículos previamente seleccionados y rechazados por el usuario para guiar el proceso de *screening*. Las evaluaciones de esta herramienta publicadas en la literatura [49] sugieren un gran ahorro de trabajo en la producción de revisiones sistemáticas a costa de una tasa de falsos negativos del 0,1 %.

Entre las revisiones analizadas en este estudio, este subgrupo presentó las puntuaciones más altas en la herramienta de valoración AMSTAR-2 (9,1/13), destacando las menciones de dos revisiones Cochrane [28, 29] (12 puntos) y un meta-análisis rápido [30] publicado en *The Lancet* (10,5 puntos). A diferencia de las revisiones de las otras categorías, que priorizaron la profundidad de la búsqueda, el uso de herramientas basadas en IA en este subgrupo estuvo motivado por la gran carga en *screening* a la que se enfrentaron los revisores: citando a Dinnes *et al.* [28], «*se necesitaba un enfoque más eficiente para procesar la creciente producción científica sobre el COVID-19*».

4. DISCUSIÓN

En este trabajo se evaluó si los potenciales beneficios de implementar IA en la elaboración de revisiones sistemáticas se han visto reflejados en las revisiones sobre el COVID-19. Encontramos que la IA se utilizó raramente, apareciendo sólo en el 0,7 % de las revisiones estudiadas, pero que se asoció significativamente con una reducción de la carga de trabajo en *screening* de los autores y la publicación en revistas con mayor factor de impacto. Ser una revisión *living* se asoció con el uso de IA, siendo los usos más comunes la optimización del *screening* (priorizando los estudios con alta probabilidad de ser relevantes) y la selección de ensayos controlados aleatorizados.

Como limitación de nuestro estudio, destacamos su baja potencia estadística debido al escaso número de revisiones que utilizaron IA. Anticipando la limitada disponibilidad de revisiones con IA, adoptamos un procedimiento de *screening* muy sensible, procesando más de 7 000 referencias bibliográficas de revisiones sistemáticas sobre el COVID-19 (combinando la consulta de expertos para la selección de palabras clave con un potente motor de búsqueda), y elegimos un ratio de 3:1 para el tamaño del grupo de control para minimizar el riesgo de errores estadísticos de tipo II. El uso de *L-OVE* como

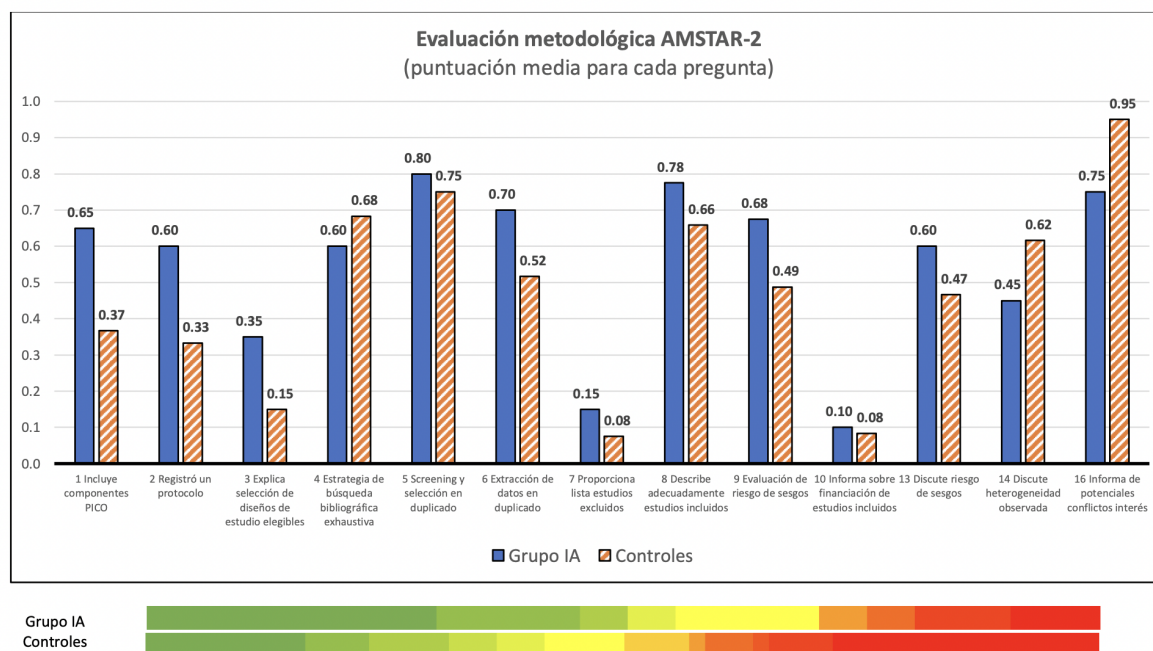


Figura 3 Resumen de la evaluación metodológica AMSTAR-2. El gráfico de la parte superior muestra las puntuaciones medias obtenidas en cada una de las preguntas evaluadas en el grupo de revisiones sistemáticas con uso de inteligencia artificial (IA) (barras azules) y en el grupo control (barras naranjas). Las líneas de colores de la parte inferior representan visualmente la heterogeneidad de resultados obtenidos en las revisiones sistemáticas de ambos grupos (el gradiente representa las puntuaciones obtenidas: rojo=4; amarillo=6,5; verde=9).

base de datos principal permitió acceder a múltiples fuentes bibliográficas relevantes y actualizadas de forma sistemática y automatizable; sin embargo, nuestra estrategia de búsqueda podría mostrar una menor sensibilidad para los informes institucionales y *whitepapers*, que a menudo no son indexados por las bases de datos tradicionales. El potencial impacto de errores puntuales en la descarga de revisiones y la exclusión de las revisiones no *open-access* de nuestro estudio es incierto; su impacto en la generalización de nuestros resultados debe interpretarse teniendo en cuenta la gran diversidad de fuentes secundarias accesibles a través de *L-OVE* y la alta accesibilidad a la investigación sobre el COVID-19 durante la pandemia. Además, el uso de las fechas de publicación como variable de emparejamiento permitió una selección de controles guiada por nuestro *script* (para minimizar el riesgo de sesgos), pero impidió el uso de otras variables de control deseables, como el tamaño o el enfoque temático de las revisiones.

Debemos resaltar que el cálculo de la carga de *screening* “por autor” en lugar de “por revisor que participa en el cribado” puede infraestimar estas variables en revisiones con equipos grandes (cuando no todos sus autores participan en el cribado). Tener mayor número de autores también podría relacionarse con la disponibilidad de recursos de los equipos, y por tanto con el acceso a asesoramiento y expertos en IA. Asimismo, los grupos con más recursos y con apoyo de expertos en IA podrían contar con mayores facilidades de acceso a revistas bien indexadas, lo que podría sesgar los análisis del factor

de impacto a favor de la IA. Por otra parte, la evaluación con AMSTAR-2 se realizó inevitablemente sin poder cegar a los evaluadores en cuanto al uso o no de IA, lo cual, dada la subjetividad de ciertos aspectos de la evaluación metodológica, podría haber influido en sus resultados. Por último, el uso de los recuentos de citas para medir el impacto de las revisiones es un abordaje imperfecto con deficiencias conocidas, como el “*citation bias*” (los resultados estadísticamente significativos suelen citarse más que otros) o la propia autoridad de los autores [50] y puede particularmente subestimar el impacto de las revisiones publicadas más recientemente.

En promedio, un equipo de 5 revisores tarda 15 meses en completar una revisión sistemática tradicional [51] con tasas de errores estimadas en torno al 10% [52]. Enfrentarse a la pandemia del COVID-19 exigía contar con revisiones sistemáticas sólidas con urgencia, ya que cualquier retraso suponía un coste tanto en términos de vidas perdidas como en daños económicos. Sin embargo, a pesar del crecimiento exponencial que han experimentado los campos de la IA y el *machine-learning* durante los últimos años, estos desempeñaron un papel sorprendentemente limitado en las revisiones sistemáticas sobre el COVID-19. Nuestras conclusiones concuerdan con las de trabajos anteriores [14] que sugieren que los beneficios que la IA puede proporcionar en la realización de revisiones sistemáticas son desconocidos para la mayoría de los revisores, y la relativa heterodoxia de sus métodos podría dificultar inicialmente su aceptación por parte de la comunidad científica. El software de código abierto, general-

mente más propenso a ser adoptado en dichos círculos, jugará un papel esencial en este aspecto.

Nuestra descripción narrativa de las revisiones incluidas en este estudio mostró que ninguna hizo uso simultáneo de más de una herramienta de IA. Un enfoque más cohesivo, que integre la IA en cada paso del proceso de revisión, ahorraría tiempo a los revisores que desearan interconectar diferentes herramientas con formatos a veces incompatibles. Las herramientas de asistencia semiautomatizadas fueron una de las áreas donde la IA mostró una mayor adopción, y la variedad de opciones de software (como *EPPI Reviewer*, ya adoptado como herramienta oficial de producción de revisiones Cochrane) fue mayor. Por el contrario, la automatización completa sólo fue empleada por *RobotSearch* (una herramienta para filtrar ensayos aleatorizados ampliamente evaluada en la literatura), lo que sugiere que la adopción de soluciones cada vez más automatizadas puede requerir la producción previa de evaluaciones más detalladas de sus costes potenciales (tanto en pérdida de artículos como en riesgo de sesgos) en balance con sus aportaciones en productividad.

Conclusión

La necesidad de automatización en las síntesis de evidencia es obvia, ya que la carga de trabajo de los revisores crece tan rápido como las ciencias biomédicas. La adopción de nuevas tecnologías puede llevar tiempo, pero aprovechar el potencial de la IA en la elaboración de revisiones sistemáticas debe ser una prioridad. En el futuro, la IA debe incorporarse a las revisiones sistemáticas como el siguiente paso para facilitar la toma de decisiones de manera más basada en la evidencia, rápida y precisa.

Material suplementario

Disponible en la publicación original.

Referencias

- [1] TJ Lasserson, J Thomas, and JPT Higgins. "Starting a review". In: *Cochrane handbook for systematic reviews of interventions*. 6.3. Cochrane, 2022. Chap. 1. URL: <https://training.cochrane.org/handbook>.
- [2] Guy Tsafnat, Paul Glasziou, Miew K. Choong, Adam Dunn, Filippo Galgani, and Enrico Coiera. "Systematic review automation technologies". In: *Systematic Reviews* 3 (2014), p. 74. DOI: 10.1186/2046-4053-3-74.
- [3] B. Nussbaumer-Streit, M. Ellen, I. Klerings, et al. "Resource use during systematic review production varies widely: a scoping review". In: *Journal of Clinical Epidemiology* 139 (2021), pp. 287–296. DOI: 10.1016/j.jclinepi.2021.05.019.
- [4] Dipender Gill, Emma H. Baker, and Andrew W. Hitchings. "We need clinical guidelines fit for a pandemic". In: *BMJ* 373 (2021), n1093. DOI: 10.1136/bmj.n1093.
- [5] J. André Knottnerus and Peter Tugwell. "Methodological challenges in studying the COVID-19 pandemic crisis". In: *Journal of Clinical Epidemiology* 121 (2020), A5–A7. DOI: 10.1016/j.jclinepi.2020.04.001.
- [6] Yanfei Li, Liujiao Cao, Ziyao Zhang, et al. "Reporting and methodological quality of COVID-19 systematic reviews needs to be improved: an evidence mapping". In: *Journal of Clinical Epidemiology* 135 (2021), pp. 17–28. DOI: 10.1016/j.jclinepi.2021.02.021.
- [7] Richard G. Jung, Pietro Di Santo, Cole Clifford, et al. "Methodological quality of COVID-19 clinical research". In: *Nature Communications* 12 (2021), p. 943. DOI: 10.1038/s41467-021-21220-5.
- [8] Andrea C. Tricco, Chantelle M. Garrity, Leah Boulos, et al. "Rapid review methods more challenging during COVID-19: commentary with a focus on 8 knowledge synthesis steps". In: *Journal of Clinical Epidemiology* 126 (2020), pp. 177–183. DOI: 10.1016/j.jclinepi.2020.06.029.
- [9] Linda Biesty, Pauline Meskell, Claire Glenton, et al. "A QuEst for speed: rapid qualitative evidence syntheses as a response to the COVID-19 pandemic". In: *Systematic Reviews* 9 (2020), p. 256. DOI: 10.1186/s13643-020-01512-5.
- [10] Carmen Amezcua-Prieto, Juan M. Fernández-Luna, Juan F. Huete-Guadix, Aurora Bueno-Cavanillas, and Khalid Saed Khan. "Artificial intelligence and automation of systematic reviews in women's health". In: *Current opinion in Obstetrics & Gynecology* 32.5 (2020), pp. 335–341. DOI: 10.1097/GCO.0000000000000643.
- [11] Alison O'Mara-Eves, James Thomas, John McNaught, Makoto Miwa, and Sophia Ananiadou. "Using text mining for study identification in systematic reviews: a systematic review of current approaches". In: *Systematic reviews* 4 (2015), p. 5. DOI: 10.1186/2046-4053-4-5.
- [12] Iain J. Marshall and Byron C. Wallace. "Toward systematic review automation: A practical guide to using machine learning tools in research synthesis". In: *Systematic Reviews* 8 (2019), p. 163. DOI: 10.1186/s13643-019-1074-9.
- [13] James Thomas, Anna Noel-Storr, Iain Marshall, et al. "Living systematic reviews: 2. Combining human and machine effort". In: *Journal of Clinical Epidemiology* 91 (2017), pp. 31–37. DOI: 10.1016/j.jclinepi.2017.08.011.
- [14] Anna Mae Scott, Connor Forbes, Justin Clark, Matt Carter, Paul Glasziou, and Zachary Munn. "Systematic review automation tools improve efficiency but lack of knowledge impedes their adoption: a survey". In: *Journal of Clinical Epidemiology* 138 (2021), pp. 80–94. DOI: 10.1016/j.jclinepi.2021.06.030.
- [15] James Thomas, Steve McDonald, Anna Noel-Storr, et al. "Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews". In: *Journal of Clinical Epidemiology* 133 (2021), pp. 140–151. DOI: 10.1016/j.jclinepi.2020.11.003.
- [16] Lena Schmidt, Babatunde K. Olorisade, Luke A. McGuinness, James Thomas, and Julian P.T. Higgins. "Data extraction methods for systematic review (semi)automation: A living systematic review". In: *F1000Research* 10 (2021), p. 401. DOI: 10.12688/f1000research.51117.1.
- [17] Siddhartha R. Jonnalagadda, Pawan Goyal, and Mark D. Huffman. "Automating data extraction in systematic reviews: A systematic review". In: *Systematic Reviews* 4 (2015), p. 78. DOI: 10.1186/s13643-015-0066-7.
- [18] Lawrence Mbuagbaw, Daeria O. Lawson, Livia Puljak, David B. Allison, and Lehana Thabane. "A tutorial on methodological studies: The what, when, how and why". In: *BMC Medical Research Methodology* 20 (2020), p. 226. DOI: 10.1186/s12874-020-01107-7.
- [19] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, et al. "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews". In: *BMJ* 372 (2021), n71. DOI: 10.1136/bmj.n71.
- [20] Juan Ramón Tercero-Hidalgo, Khalid Saeed Khan, Aurora Bueno-Cavanillas, et al. "Covid-19 systematic evidence synthesis with Artificial Intelligence: A Review of Reviews". In: *Open Science Forum Registries* (2021). DOI: 10.17605/OSF.IO/H5DAW.
- [21] Juan Ramón Tercero-Hidalgo, Khalid Saeed Khan, Aurora Bueno-Cavanillas, et al. "COVID-19 evidence syntheses with artificial intelligence: an empirical study of systematic reviews". In: *Dryad Dataset* (2021). DOI: 10.5061/dryad.9kd51c5j6.
- [22] Gabriel Rada, Francisca Verdugo-Paiva, Camila Ávila, et al. "Evidence synthesis relevant to COVID-19: a protocol for multiple

- systematic reviews and overviews of systematic reviews". eng. In: *Medwave* 20.3 (2020), e7868. DOI: [10.5867/medwave.2020.03.7867](https://doi.org/10.5867/medwave.2020.03.7867).
- [23] Kerry Dhakal. "Unpaywall". In: *Journal of the Medical Library Association: JMLA* 107.2 (2019), pp. 286–288. DOI: [10.5195/jmla.2019.650](https://doi.org/10.5195/jmla.2019.650).
- [24] Beverley J. Shea, Barnaby C. Reeves, George Wells, et al. "AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both". In: *BMJ* 358 (2017), j4008. DOI: [10.1136/bmj.j4008](https://doi.org/10.1136/bmj.j4008).
- [25] Jessica J. Bartoszko, Reed A.C. Siemieniuk, Elena Kum, et al. "Prophylaxis against covid-19: Living systematic review and network meta-analysis". In: *BMJ* 373 (2021), n949. DOI: [10.1136/bmj.n949](https://doi.org/10.1136/bmj.n949).
- [26] Reed A C Siemieniuk, Jessica J Bartoszko, Long Ge, et al. "Drug treatments for covid-19: Living systematic review and network meta-analysis". In: *BMJ* 370 (2020), p. m2980. DOI: [10.1136/bmj.m2980](https://doi.org/10.1136/bmj.m2980).
- [27] Laure Wynants, Ben Van Calster, Gary S. Collins, et al. "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal". In: *BMJ* 369 (2020), p. m1328. DOI: [10.1136/bmj.m1328](https://doi.org/10.1136/bmj.m1328).
- [28] Jacqueline Dinnes, Jonathan J. Deeks, Sarah Berhane, et al. "Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection". In: *Cochrane Database of Systematic Reviews* 3 (2021), p. CD013705. DOI: [10.1002/14651858.CD013705.pub2](https://doi.org/10.1002/14651858.CD013705.pub2).
- [29] Thomas Struyf, Jonathan J. Deeks, Jacqueline Dinnes, et al. "Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19". In: *Cochrane Database of Systematic Reviews* 2 (2021), p. CD013665. DOI: [10.1002/14651858.CD013665.pub2](https://doi.org/10.1002/14651858.CD013665.pub2).
- [30] Derek K. Chu, Elie A. Akl, Stephanie Duda, et al. "Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis". In: *The Lancet* 395.10242 (2020), pp. 1973–1987. DOI: [10.1016/S0140-6736\(20\)31142-9](https://doi.org/10.1016/S0140-6736(20)31142-9).
- [31] Sravanthi Parasa, Madhav Desai, Viveksandeep Thoguluva Chandrasekar, et al. "Prevalence of Gastrointestinal Symptoms and Fecal Viral Shedding in Patients with Coronavirus Disease 2019: A Systematic Review and Meta-analysis". In: *JAMA Network Open* 3.6 (2020), e2011335. DOI: [10.1001/jamanetworkopen.2020.11335](https://doi.org/10.1001/jamanetworkopen.2020.11335).
- [32] Nazar Zaki, Elfadil Abdalla Mohamed, Sahar Ibrahim, and Gulfaraz Khan. "The influence of comorbidity on the severity of COVID-19 disease: systematic review and analysis". In: *medRxiv* (2020), p. 2020.06.18.20134478. DOI: [10.1101/2020.06.18.20134478](https://doi.org/10.1101/2020.06.18.20134478).
- [33] Nazar Zaki and Elfadil A. Mohamed. "The estimations of the COVID-19 incubation period: A scoping reviews of the literature". In: *Journal of Infection and Public Health* 14.5 (2021), pp. 638–646. DOI: [10.1016/j.jiph.2021.01.019](https://doi.org/10.1016/j.jiph.2021.01.019).
- [34] Matthew Michelson, Tiffany Chow, Neil Martin, Mike Ross, Amelia Tee, and Steven Minton. "Ocular toxicity and Hydroxychloroquine: A Rapid Meta-Analysis". In: *medRxiv* (2020), p. 2020.04.28.20083378. DOI: [10.1101/2020.04.28.20083378](https://doi.org/10.1101/2020.04.28.20083378).
- [35] Caitlin Daley, Megan Fydenkevez, and Shari Ackerman-Morris. "A Systematic Review of the Incubation Period of SARS-CoV-2: The Effects of Age, Biological Sex, and Location on Incubation Period". In: *medRxiv* (2020), p. 2020.12.23.20248790. DOI: [10.1101/2020.12.23.20248790](https://doi.org/10.1101/2020.12.23.20248790).
- [36] Robert Robinson, Vidhya Prakash, Raad Al Tamimi, et al. "Impact of remdesivir on 28 day mortality in hospitalized patients with COVID-19: February 2021 Meta-analysis". In: *medRxiv* (2021), p. 2021.03.04.21252903. DOI: [10.1101/2021.03.04.21252903](https://doi.org/10.1101/2021.03.04.21252903).
- [37] Robert Robinson, Vidhya Prakash, Raad Al Tamimi, Nour Albast, and Basma Al-Bast. "Impact of systemic corticosteroids on hospitalized patients with COVID-19: January 2021 Meta-analysis of randomized controlled trials". In: *medRxiv* (2021), p. 2021.02.03.21251065. DOI: [10.1101/2021.02.03.21251065](https://doi.org/10.1101/2021.02.03.21251065).
- [38] Ariel Izcovich, Reed Alexander Siemieniuk, Jessica Julia Bartoszko, et al. "Adverse effects of remdesivir, hydroxychloroquine and lopinavir/ritonavir when used for COVID-19: systematic review and meta-analysis of randomised trials". In: *BMJ Open* 12.3 (2022), e048502. DOI: [10.1136/bmjopen-2020-048502](https://doi.org/10.1136/bmjopen-2020-048502).
- [39] Dena Zeraatkar, Ellen Cusano, Juan Pablo Díaz Martínez, et al. "Use of tocilizumab and sarilumab alone or in combination with corticosteroids for covid-19: systematic review and network meta-analysis". In: *BMJ Medicine* 1.1 (2022), e000036. DOI: [10.1101/2021.07.05.21259867](https://doi.org/10.1101/2021.07.05.21259867).
- [40] Kimia Honarmand, Jeremy Penn, Arnav Agarwal, et al. "Clinical trials in COVID-19 management & prevention: A meta-epidemiological study examining methodological quality". In: *Journal of Clinical Epidemiology* 139 (2021), pp. 68–79. DOI: [10.1016/j.jclinepi.2021.07.002](https://doi.org/10.1016/j.jclinepi.2021.07.002).
- [41] Russell Viner, Simon Russell, Rosella Saulle, et al. "Impacts of school closures on physical and mental health of children and young people: a systematic review". In: *medRxiv* (2021), p. 2021.02.10.21251526. DOI: [10.1101/2021.02.10.21251526](https://doi.org/10.1101/2021.02.10.21251526).
- [42] J. Wilson, G. Carson, S. Fitzgerald, et al. "Are medical procedures that induce coughing or involve respiratory suctioning associated with increased generation of aerosols and risk of SARS-CoV-2 infection? A rapid systematic review". In: *Journal of Hospital Infection* 116 (2021), pp. 37–46. DOI: [10.1016/j.jhin.2021.06.011](https://doi.org/10.1016/j.jhin.2021.06.011).
- [43] Rebecca Elmore, Lena Schmidt, Juleen Lam, et al. "Risk and Protective Factors in the COVID-19 Pandemic: A Rapid Evidence Map". In: *Frontiers in Public Health* 8 (2020), p. 582205. DOI: [10.3389/fpubh.2020.582205](https://doi.org/10.3389/fpubh.2020.582205).
- [44] Hadeel Alkofide, Abdullah Almohaizeie, Sara Almuhaiani, Bashayer Alotaibi, and Khalid M. Alkharfy. "Tocilizumab and Systemic Corticosteroids in the Management of Patients with COVID-19: A Systematic Review and Meta-Analysis". In: *International Journal of Infectious Diseases* 110 (2021), pp. 320–329. DOI: [10.1016/j.ijid.2021.07.021](https://doi.org/10.1016/j.ijid.2021.07.021).
- [45] Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, et al. "CORD-19: The Covid-19 Open Research Dataset". In: *ArXiv* (2020), p. 2004.10706v4. DOI: [10.48550/arXiv.2004.10706](https://doi.org/10.48550/arXiv.2004.10706).
- [46] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, et al. "BioBERT: A pre-trained biomedical language representation model for biomedical text mining". In: *Bioinformatics* 36.4 (2020), pp. 1234–1240. DOI: [10.1093/bioinformatics/btz682](https://doi.org/10.1093/bioinformatics/btz682).
- [47] Iain J. Marshall, Anna Noel-Storr, Joël Kuiper, James Thomas, and Byron C. Wallace. "Machine learning for identifying Randomized Controlled Trials: An evaluation and practitioner's guide". In: *Research Synthesis Methods* 9.4 (2018), pp. 602–614. DOI: [10.1002/jrsm.1287](https://doi.org/10.1002/jrsm.1287).
- [48] Brian E. Howard, Jason Phillips, Arpit Tandon, et al. "SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation". In: *Environment International* 138 (2020), p. 105623. DOI: [10.1016/j.envint.2020.105623](https://doi.org/10.1016/j.envint.2020.105623).
- [49] Allison Gates, Cydney Johnson, and Lisa Hartling. "Technology-assisted title and abstract screening for systematic reviews: A retrospective evaluation of the Abstrackr machine learning tool". In: *Systematic Reviews* 7 (2018), p. 45. DOI: [10.1186/s13643-018-0707-8](https://doi.org/10.1186/s13643-018-0707-8).
- [50] Miriam J.E. Urlings, Bram Duyx, Gerard M.H. Swaen, Lex M. Bouter, and Maurice P. Zeegers. "Citation bias and other determinants of citation in biomedical research: findings from six citation networks". In: *Journal of Clinical Epidemiology* 132 (2021), pp. 71–78. DOI: [10.1016/j.jclinepi.2020.11.019](https://doi.org/10.1016/j.jclinepi.2020.11.019).
- [51] Rohit Borah, Andrew W. Brown, Patrice L. Capers, and Kathryn A. Kaiser. "Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry". In: *BMJ Open* 7 (2017), e012545. DOI: [10.1136/bmjopen-2016-012545](https://doi.org/10.1136/bmjopen-2016-012545).
- [52] Zhen Wang, Tarek Nayfeh, Jennifer Tetzlaff, Peter O'Brien, and Mohammad Hassan Murad. "Error rates of human reviewers during abstract screening in systematic reviews". In: *PLoS ONE* 15.1 (2020), e0227742. DOI: [10.1371/journal.pone.0227742](https://doi.org/10.1371/journal.pone.0227742).